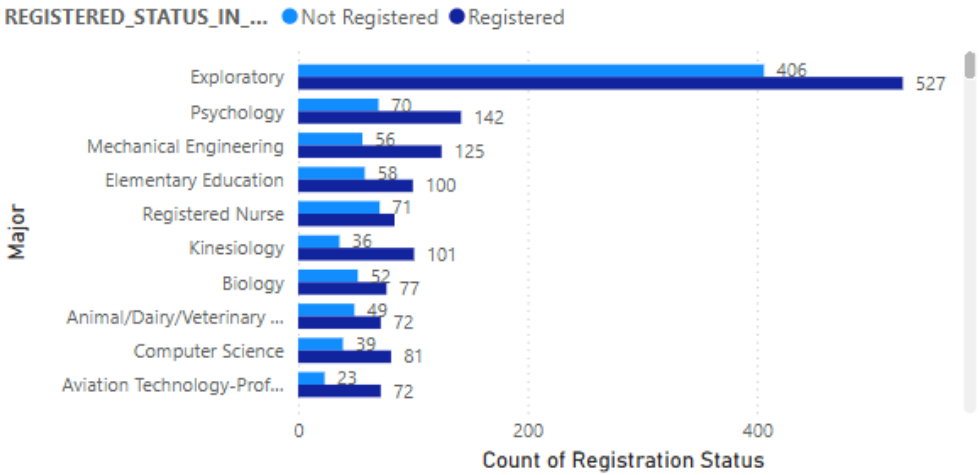
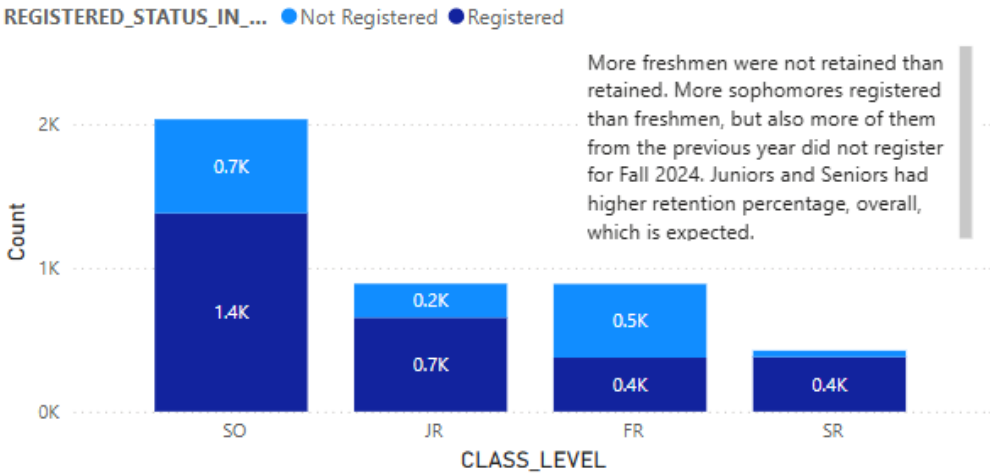
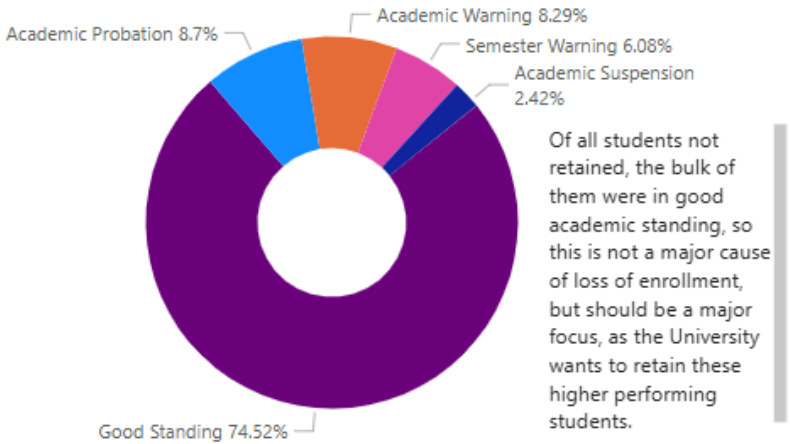
Fall 2024 Retention by Major



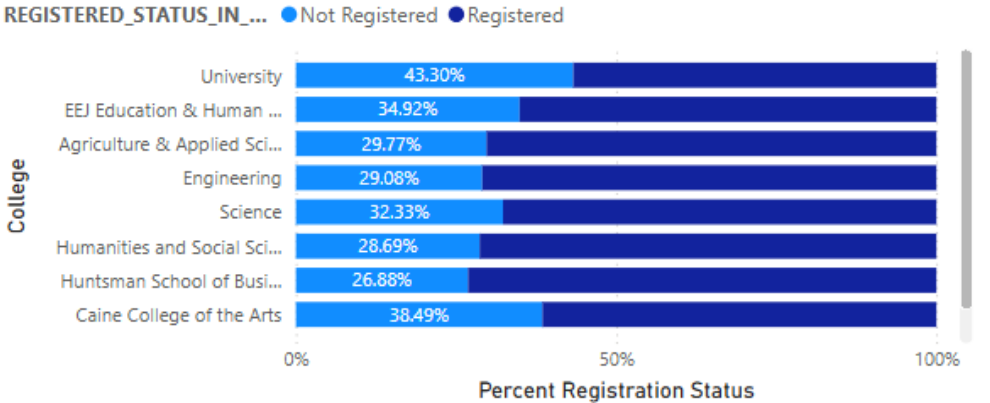
Retention by Class Level



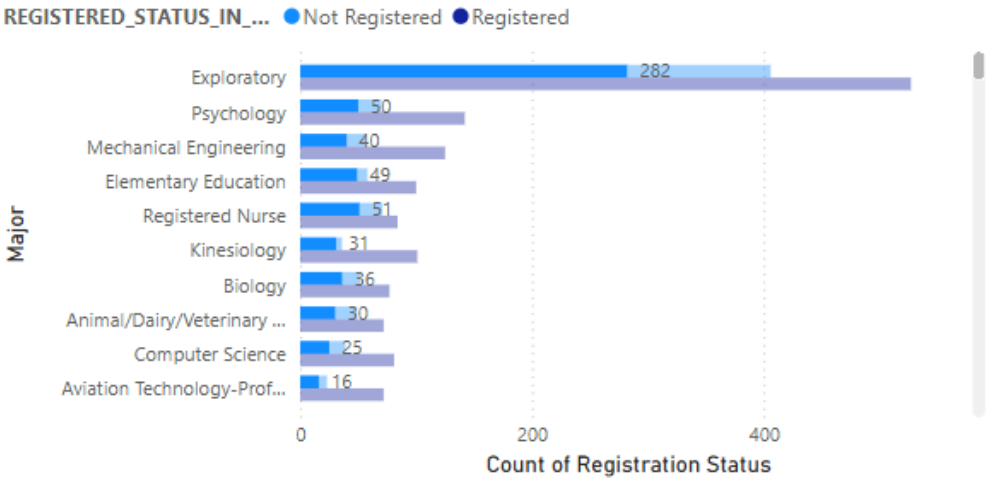
Not Retained by Academic Standing



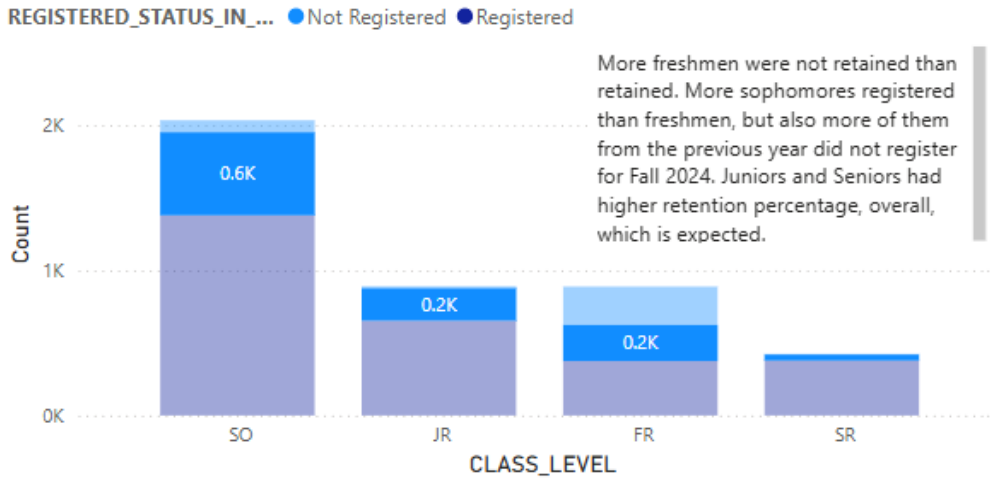
Percentage of Students Not Retained by College



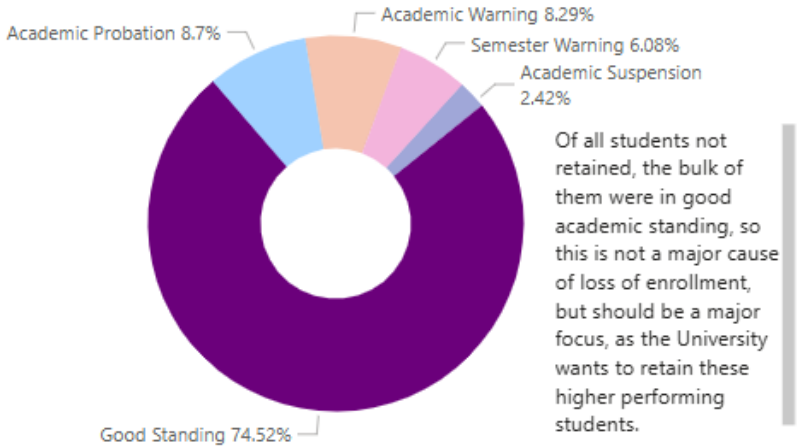
Fall 2024 Retention by Major



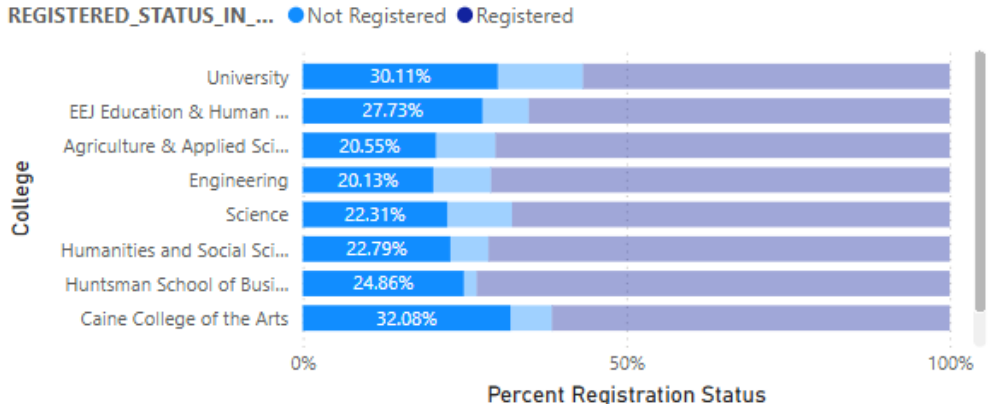
Retention by Class Level



Not Retained by Academic Standing



Percentage of Students Not Retained by College



```
LogisticRegression()
```

```
# Evaluate model
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy: {:.2f}%".format(accuracy * 100))
```

```
Accuracy: 69.46%
```

```
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
```

```
Confusion Matrix:
 [[ 71 216]
 [ 43 518]]

Classification Report:
              precision    recall  f1-score   support

           0       0.62      0.25      0.35       287
           1       0.71      0.92      0.80       561

    accuracy                           0.69       848
   macro avg       0.66      0.59      0.58       848
weighted avg       0.68      0.69      0.65       848
```

Analysis of Factors Contributing to Student Retention

Kristen Sohm
kristensohm@gmail.com

Abstract:

This analysis examines factors contributing to retention at a university. Of the sample of 4,258 students who registered in Fall 2023, 34% did not register in Fall 2024. Findings suggest Sophomores with undeclared majors are at higher risk for not returning to the university, which highlights need for more support for students around major declaration.

Analysis:

Using Python in JupyterLab, I cleaned the dataset of 16 attributes and 4,258 instances, dealt with empty fields, one-hot encoded all of the categorical variables, and trained and tested a logistic regression model to see how well these variables predict retention, which was at 69.46% accuracy. As seen in the attached "RetentionClassificationReport.png", this model is better at predicting the positive class [1] (Predicting which students will register) than predicting which students will not register [0]. This model could use more variables to increase accuracy, but the main issue is what is causing students not to re-register?

I performed a feature importance analysis for the logistic regression model to isolate factors that best predict retention. For this model, the most important factors relate to major, class level, college, and academic standing.

So, I graphed retention (whether students registered or not) for each of these attributes using Power BI. Attached is a report summary of visualizations, "RetentionFactorsVisualReport.png", of how these four factors contribute to retention. The graph 'Not Retained by Academic Standing' shows academic standing is not the major reason students decide not to register. The bulk of students that were not retained were in good academic standing.

However, the University would be most interested in these high achieving students and would like to retain more of them. So, attached is "StudentsNotRetainedButInGoodStanding.png", a static view of an interactive report where I selected those students that were not retained but had good academic standing. The report highlights the portions of all other graphs that were due to these specific students that were not retained. This information could be used to tailor data collection toward learning more about the academically strong students that decide not to continue at the university and thus be able to focus retention efforts in those key areas.

For example, a large portion of the students in 'Exploratory Major' were those students that did not come back but were in good standing. We can tailor and analyze exit surveys to see what other factors cause students to leave and focus energies in those areas to improve student retention. Particularly, the Exploratory Major, is the largest Major category by far of both registered students and where students most commonly decide not to register.

Retention by class level shows a larger proportion of freshmen did not do well academically. Sophomores did proportionately better academically, but also had a larger proportion of students that did not register for Fall 2024.

The percentage of students not retained by each college shows that Caine College of the Arts has a higher rate of not registering than University College. Future analysis could investigate factors contributing to lower retention rates for individual colleges. Surveys could ask questions about specific majors to target reasons for leaving the University, which may be different than for the Exploratory Major.

Now we know most of the students that did not register were in good standing, in an Exploratory Major, and were Sophomores. The recommendation is to increase decision support for Sophomores in good academic standing with undeclared majors. Future analysis could explore factors related to major declaration decisions.